Randomized Algorithms

南京大学

尹一通

Mixing Time

Markov chain: $\mathfrak{M} = (\Omega, P)$

stationary distribution: π

$p_x^{(t)}$: distribution at time t when initial state is x

$$\Delta_x(t) = \|p_x^{(t)} - \pi\|_{TV} \qquad \Delta(t) = \max_{x \in \Omega} \Delta_x(t)$$

$$\tau_x(\epsilon) = \min\{t \mid \Delta_x(t) \le \epsilon\} \qquad \tau(\epsilon) = \max_{x \in \Omega} \tau_x(\epsilon)$$

- mixing time: $\tau_{\text{mix}} = \tau(1/2\mathrm{e})$

  rapid mixing: $\tau_{\text{mix}} = (\log |\Omega|)^{O(1)}$

$$\Delta(k \cdot \tau_{\text{mix}}) \le \mathrm{e}^{-k} \quad \text{and} \quad \tau(\epsilon) \le \tau_{\text{mix}} \cdot \left\lceil \ln \frac{1}{\epsilon} \right\rceil$$

Coupling of Markov Chains

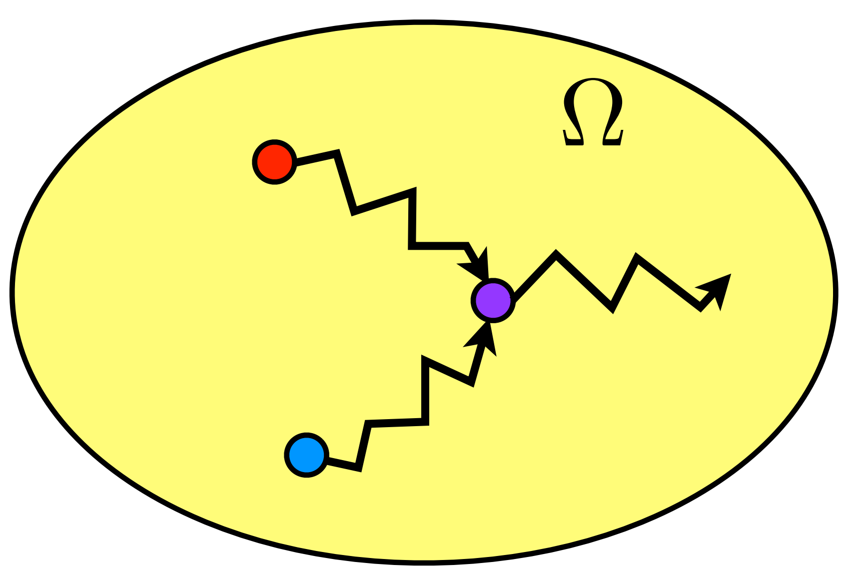
a coupling of $\mathfrak{M} = (\Omega, P)$ is a Markov chain (X_t, Y_t) of state space $\Omega \times \Omega$ such that:

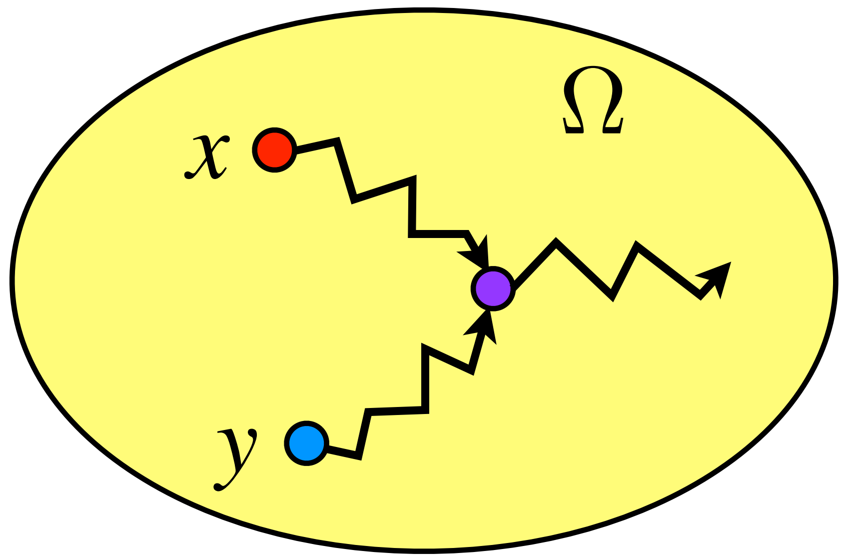
- both are faithful copies of the chain

$$\Pr[X_{t+1} = y \mid X_t = x] = \Pr[Y_{t+1} = y \mid Y_t = x] = P(x, y)$$

- once collides, always makes identical moves

$$X_t = Y_t \Rightarrow X_{t+1} = Y_{t+1}$$





Markov Chain Coupling Lemma:

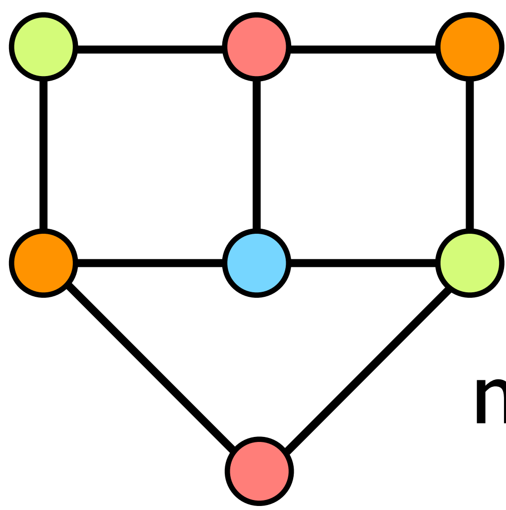
(X_t, Y_t) is a coupling of $\mathfrak{M} = (\Omega, P)$ ⇨

$$\Delta(t) \le \max_{x,y\in\Omega} \Pr[X_t \neq Y_t \mid X_0 = x, Y_0 = y]$$

$$\max_{x,y\in\Omega} \Pr[X_t \neq Y_t \mid X_0 = x, Y_0 = y] \le \epsilon \quad \Rightarrow \quad \tau(\epsilon) \le t$$

Graph Coloring

$G(V,E)$



max degree Δ

proper q-coloring $f : V \to [q]$

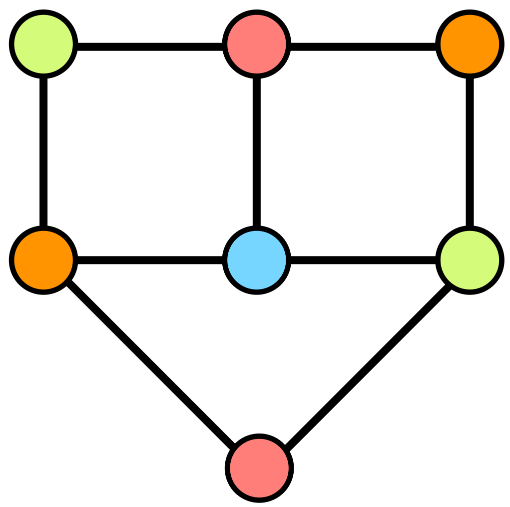
$$\forall uv \in E, \quad f(u) \neq f(v)$$

decision: Is G q-colorable?

- $q<\Delta$: NP-hard;
- $q=\Delta$: q-colorable unless G has $(\Delta+1)$-clique or G is an odd cycle; (Brooks Theorem)
- $q\geq\Delta+1$: always q-colorable and the q-coloring can be found by a greedy algorithm;

sampling: sample a uniform random proper q-coloring

counting: How many proper q-colorings for G?



$G(V,E)$ of max degree Δ

proper q-coloring with $q \geq \alpha\Delta + \beta$

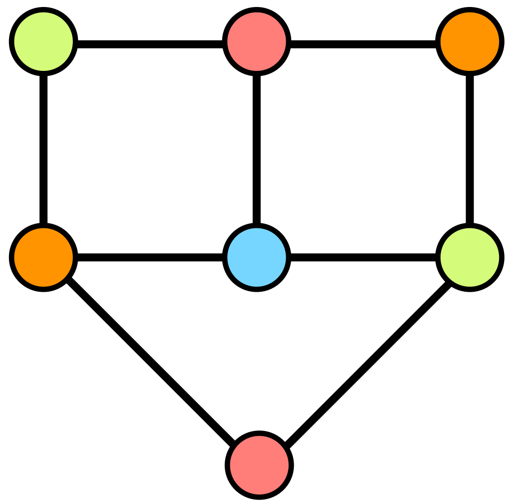
sampling: sample a uniform random proper q-coloring

Markov Chain (Glauber dynamics):

at each step:

- randomly pick a vertex $v \in V$ and a color $c \in [q]$;
- change the color of v to c if it is proper;

$q \geq \Delta + 2$ ⇨ aperiodic; irreducible; uniform stationary distribution;



$G(V,E)$ of max degree Δ

proper q-coloring with $q \geq \alpha\Delta + \beta$

sampling: sample a uniform random proper q-coloring

Markov Chain (Glauber dynamics):

at each step:

- randomly pick a vertex $v \in V$ and a color $c \in [q]$;
- change the color of v to c if it is proper;

Conjecture

$q \geq \Delta + 2$ ⇨ Glauber dynamics is rapid mixing at each step:

- randomly pick a vertex $v \in V$ and a color $c \in [q]$;
- change the color of v to c if it is proper;

Theorem (Jerrum 1995)

$q \geq 4\Delta + 1$ ⇨ rapid mixing

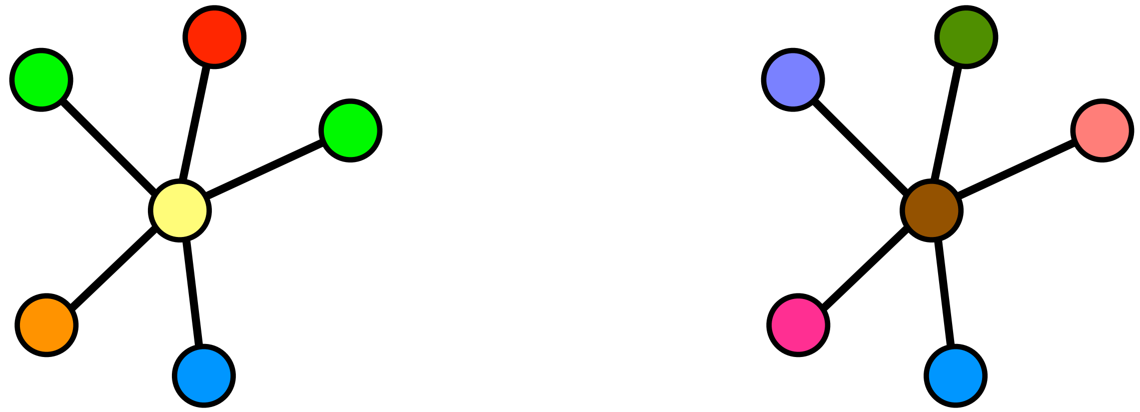
coupling rule: $(X_t, Y_t) \in \Omega \times \Omega$

at each step, choose the same $v \in V$ and $c \in [q]$

X_{t+1}	Y_{t+1}
changed	changed
unchanged	changed
changed	unchanged
unchanged	unchanged

$d_t = d(X_t, Y_t)$: Hamming distance

- good move: distance decreases by 1
- bad move: distance increases by 1
- neutral move: distance unchanged



at each step, choose the same $v \in V$ and $c \in [q]$

X_{t+1}	Y_{t+1}
changed	changed
unchanged	changed
changed	unchanged
unchanged	unchanged

$d_t = d(X_t, Y_t)$: Hamming distance

- good move: distance decreases by 1
- bad move: distance increases by 1
- neutral move: distance unchanged

# of good moves: $\geq d_t(q - 2\Delta)$

v is a disagreeing vertex, c is not in both neighborhoods

# of bad moves: $\leq 2d_t\Delta$

v is a neighbor of disagreeing vertex, c is one of the two colors at each step, choose the same $v \in V$ and $c \in [q]$

$d_t = d(X_t, Y_t)$: Hamming distance

- good move: distance decreases by 1
- bad move: distance increases by 1
- neutral move: distance unchanged

\# of good moves: $\geq d_t(q - 2\Delta)$

\# of bad moves: $\leq 2d_t\Delta$

$$\mathbf{E}[d_{t+1} \mid d_t] \leq d_t - \frac{d_t(q-2\Delta)}{qn} + \frac{2d_t\Delta}{qn} = d_t\left(1 - \frac{q-4\Delta}{qn}\right)$$

$$\Rightarrow \quad \mathbf{E}[d_{t+1} \mid d_0] \leq \left(1 - \frac{q-4\Delta}{qn}\right)\mathbf{E}[d_t \mid d_0]$$

$$\leq d_0\left(1 - \frac{q-4\Delta}{qn}\right)^{(t+1)} \leq n\left(1 - \frac{1}{qn}\right)^{t+1}$$

when $q \geq 4\Delta + 1$

at each step, choose the same $v \in V$ and $c \in [q]$

$q \geq 4\Delta + 1$ ⇨ $\mathbf{E}[d_t \mid d_0] \leq n\left(1 - \frac{1}{qn}\right)^t$

Markov Chain coupling lemma:

$$\Delta(t) \leq \max_{x,y\in\Omega} \Pr[X_t \neq Y_t \mid X_0 = x, Y_0 = y]$$

$$\leq \max_{x,y\in\Omega} \Pr[d_t \geq 1 \mid X_0 = x, Y_0 = y]$$

$$\leq \max_{x,y\in\Omega} \mathbf{E}[d_t \mid d(x,y)]$$ (Markov inequality)

$$\leq n\left(1 - \frac{1}{qn}\right)^t = \epsilon$$

$$\tau(\epsilon) = qn(\ln n + \ln \tfrac{1}{\epsilon}) \qquad \tau_{\mathrm{mix}} = O(qn \log n)$$

Mixing

- Why should a Markov chain be rapidly mixing?
- Why should a random walk on a regular graph be rapidly mixing?

initial distribution q

the decreasing rate of $\|qP^t - \pi\|_1$

Spectral Decomposition

Spectral Theorem

P : symmetric $n \times n$ matrix

eigenvalues : $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$

⇨ the corresponding eigenvectors $v_1, v_2, \ldots, v_n$ are orthonormal

$\forall q \in \mathbb{R}^n \qquad q = \sum_{i=1}^{n} c_i v_i$ where $c_i = q^T v_i$

$$qP = \sum_{i=1}^{n} c_i v_i P = \sum_{i=1}^{n} c_i \lambda_i v_i$$

Mixing of Symmetric Chain

$\mathfrak{M} = ([n], P)$ P is symmetric stationary $\pi = \left(\frac{1}{n}, \ldots, \frac{1}{n}\right)$

eigenvalues : $1 = \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ (Perron-Frobenius)

orthonormal eigenbasis : $v_1, v_2, \ldots, v_n$

$q \in [0,1]^n$ is a distribution $\|q\|_1 = 1$

$\lambda_1 = 1$, $\mathbf{1}P = \mathbf{1}$ ⇒ $v_1 = \frac{\mathbf{1}}{\|\mathbf{1}\|_2} = \left(\frac{1}{\sqrt{n}}, \ldots, \frac{1}{\sqrt{n}}\right)$

⇒ $c_1 = q^T v_1 = \frac{1}{\sqrt{n}}$

$c_1 v_1 = \left(\frac{1}{n}, \ldots, \frac{1}{n}\right) = \pi$

$$q = \sum_{i=1}^{n} c_i v_i = \pi + \sum_{i=2}^{n} c_i v_i \quad \text{where} \quad c_i = q^T v_i$$

$$qP^t = \pi P^t + \sum_{i=2}^{n} c_i v_i P^t = \pi + \sum_{i=2}^{n} c_i \lambda_i^t v_i$$

$\mathfrak{M} = ([n], P)$ P is symmetric stationary $\pi = \left(\frac{1}{n}, \ldots, \frac{1}{n}\right)$

eigenvalues : $1 = \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$

orthonormal eigenbasis : $v_1, v_2, \ldots, v_n$

$q \in [0,1]^n$ is a distribution where $c_i = q^T v_i$

$$qP^t = \pi P^t + \sum_{i=2}^{n} c_i v_i P^t = \pi + \sum_{i=2}^{n} c_i \lambda_i^t v_i$$

$\mathfrak{M} = ([n], P)$ P is symmetric stationary $\pi = \left(\frac{1}{n}, \ldots, \frac{1}{n}\right)$

eigenvalues : $1 = \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$

orthonormal eigenbasis : $v_1, v_2, \ldots, v_n$

$q \in [0,1]^n$ is a distribution where $c_i = q^T v_i$

$$\|qP^t - \pi\|_1 = \left\| \sum_{i=2}^{n} c_i \lambda_i^t v_i \right\|_1 \leq \sqrt{n} \left\| \sum_{i=2}^{n} c_i \lambda_i^t v_i \right\|_2$$ (Cauchy-Schwarz)

$$= \sqrt{n} \sqrt{\sum_{i=2}^{n} c_i^2 \lambda_i^{2t}} \leq \sqrt{n} \lambda_{\max}^t \sqrt{\sum_{i=2}^{n} c_i^2}$$

define $\lambda_{\max} \triangleq \max\{|\lambda_2|, |\lambda_n|\}$

$$\leq \sqrt{n} \lambda_{\max}^t \|q\|_2 \leq \sqrt{n} \lambda_{\max}^t$$

$$\Delta(t) \leq \frac{\sqrt{n}}{2} \lambda_{\max}^t \leq \frac{\sqrt{n}}{2} \mathrm{e}^{-t(1-\lambda_{\max})} \Rightarrow \tau(\epsilon) \leq \frac{\frac{1}{2}\ln n + \ln \frac{1}{2\epsilon}}{1 - \lambda_{\max}}$$

$\mathfrak{M} = (\Omega, P)$ stationary distribution: π

$p_x^{(t)}$: distribution at time t when initial state is x

$$\Delta_x(t) = \|p_x^{(t)} - \pi\|_{TV} \qquad \Delta(t) = \max_{x \in \Omega} \Delta_x(t)$$

$$\tau_x(\epsilon) = \min\{t \mid \Delta_x(t) \leq \epsilon\} \qquad \tau(\epsilon) = \max_{x \in \Omega} \tau_x(\epsilon)$$

Theorem

P is symmetric, with eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$

Let $\lambda_{\max} = \max\{|\lambda_2|, |\lambda_n|\}$

$$\tau(\epsilon) \leq \frac{\frac{1}{2}\ln n + \ln \frac{1}{2\epsilon}}{1 - \lambda_{\max}}$$

Lazy Random Walk

- undirected d-regular graph $G(V, E)$
- lazy random walk: flip a coin to decide whether to stay

$$P(u,v) = \begin{cases} \frac{1}{2} & u = v \\ \frac{1}{2d} & u \sim v \\ 0 & \text{otherwise} \end{cases}$$

adjacency matrix A $\quad d = \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq -d$

$P = \frac{1}{2}(I + \frac{1}{d}A)$ is symmetric $\quad \nu_i = \frac{1}{2}(1 + \frac{1}{d}\lambda_i)$

eigenvalues: $\quad 1 = \nu_1 \geq \nu_2 \geq \cdots \geq \nu_n \geq 0$

adjacency matrix A $\quad d = \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq -d$

$P = \frac{1}{2}(I + \frac{1}{d}A)$ is symmetric $\quad \nu_i = \frac{1}{2}(1 + \frac{1}{d}\lambda_i)$

eigenvalues: $\quad 1 = \nu_1 \geq \nu_2 \geq \cdots \geq \nu_n \geq 0$

$$\nu_{\max} = \nu_2$$

Theorem

P is symmetric, with eigenvalues $\nu_1 \geq \nu_2 \geq \cdots \geq \nu_n$

Let $\nu_{\max} = \max\{|\nu_2|, |\nu_n|\}$

$$\tau(\epsilon) \leq \frac{\frac{1}{2}\ln n + \ln \frac{1}{2\epsilon}}{1 - \nu_{\max}}$$

Graph Spectrum

d-regular undirected graph $G(V,E)$

adjacency matrix A

eigenvalues: $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ ← graph spectrum

Theorem

Lazy random walk on d-regular graph with spectrum $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ has mixing rate

$$\tau(\epsilon) \leq \frac{d(\ln n + \ln \frac{1}{2\epsilon})}{d - \lambda_2}$$

Graph Spectrum

d-regular undirected graph $G(V,E)$

graph spectrum : $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$

1. $\forall i,\ |\lambda_i| \leq d.$
2. $\lambda_1 = d.$
3. Connected $\Leftrightarrow \lambda_1 > \lambda_2.$

d-regular undirected graph $G(V,E)$

graph spectrum : $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$

1. $\forall i,\ |\lambda_i| \leq d.$
2. $\lambda_1 = d.$
3. Connected $\Leftrightarrow \lambda_1 > \lambda_2.$

suppose $Av = \lambda v$ $\quad v_i$ has the max $|v_i|$

$$\sum_j A_{ij} v_j = \lambda v_i$$

$$|\lambda||v_i| = \left|\sum_j A_{ij} v_j\right| \leq \sum_j A_{ij}|v_j| \leq |v_i| \sum_j A_{ij} \leq d|v_i|$$

d-regular undirected graph $G(V,E)$

graph spectrum : $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$

1. $\forall i, \ |\lambda_i| \leq d.$
2. $\lambda_1 = d.$
3. Connected $\Leftrightarrow \lambda_1 > \lambda_2.$

suppose $Av = dv$ $\quad v_i$ has the max $|v_i|$

$$\left.\begin{aligned} \sum_j A_{ij} v_j &= d v_i \\ \sum_j A_{ij} &= d \end{aligned}\right\} \Rightarrow v_i = v_j \text{ for } i \sim j \quad A_{ij} > 0$$

G connected $\Rightarrow$ all v_i are equal

λ_1 has multiplicity 1

d-regular undirected graph $G(V,E)$

graph spectrum : $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$

1. $\forall i, \ |\lambda_i| \leq d.$
2. $\lambda_1 = d.$
3. $\text{Connected} \Leftrightarrow \lambda_1 > \lambda_2.$

spectral gap : $d - \lambda_2 = \lambda_1 - \lambda_2$

Theorem

$$\tau(\epsilon) \leq \frac{d(\ln n + \ln \frac{1}{2\epsilon})}{d - \lambda_2}$$

Expander graphs

Wikipedia:

"Expander graphs have found extensive applications in ***computer science****, in designing algorithms, error correcting codes, extractors, pseudorandom generators, sorting networks and robust computer networks. They have also been used in proofs of many important results in* ***computational complexity theory****, such as SL=L and the PCP theorem. In* ***cryptography*** *too, expander graphs are used to construct hash functions."*

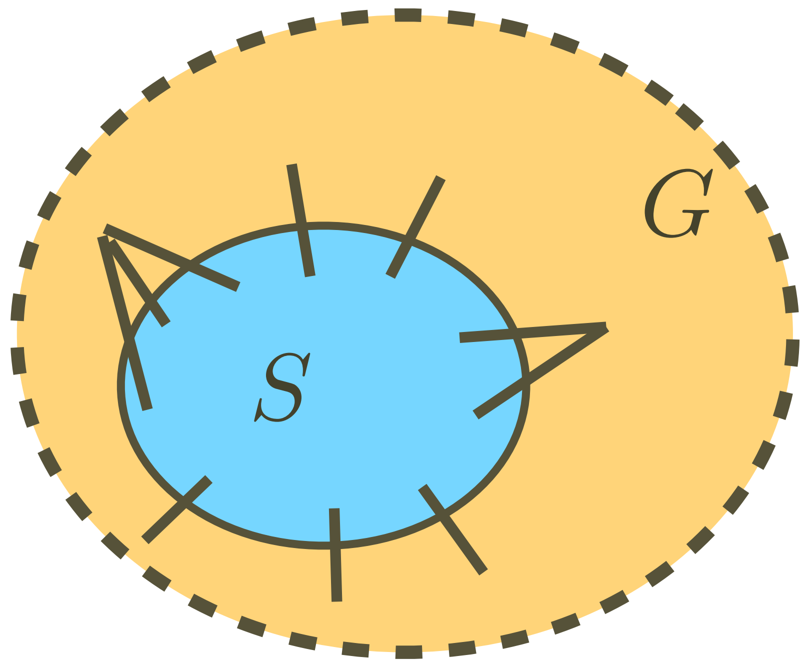
Expansion

undirected $G(V, E)$

$$E(S,T) = \{uv \in E \mid u \in S, v \in T\}$$

edge boundary

$$\partial S = E(S, \bar{S})$$



expansion ratio

$$\phi(G) = \min_{\substack{S \subset V \\ |S| \leq \frac{n}{2}}} \frac{|\partial S|}{|S|}$$

Expander Graph

$$\phi(G) = \min_{\substack{S \subset V \\ |S| \leq \frac{n}{2}}} \frac{|\partial S|}{|S|}$$

Expander graphs (combinatorial definition):

d-regular graphs with constant degree d

and constant expansion ratio $\phi(G)$.

- sparse;
- "expanding" (well connected);

“A Magical Graph!”

- Existence ?
  - random graph is an expander w.h.p.
- Computation ?
  - co-NP-complete